Apply control-system theory to analyze closed-loop systems

Dynamic signal analyzers (DSAs), which analyze signals in both the time and frequency domains, aid you in taking control-system measurements and in performing other steps in system development, such as analysis, modeling, and design. Part 1 of this 3-part article presents an overview of classical linear control theory. Part 2 will examine both old and new methods of taking the actual measurements, and part 3 will explore the expanded role of DSAs in the control-system design process.

Steve Asbjornsen and Owen Brown,
Hewlett-Packard Co

Control-system development is largely a study of the operating characteristics of devices and the ways in which the devices interact when combined in a system. You can apply linear control-system theory to analyze any system that uses deliberate guidance or manipulation to achieve a specific value for some variable, whether it's electrical, mechanical, or biological. These systems can range from configurations as simple as Watt's flyball governor to networks whose analysis requires calculations that only a computer can handle adequately. Examples of familiar control systems include motor-speed controls, pacemakers, voltage regulators, switching power supplies, and phase-locked loops.

The most fundamental distinction in control theory is the classification of systems. Without exception, systems fall into two categories: open-loop and closed-loop systems. Open-loop control systems (**Fig 1a**) are ones that use a controller that has no feedback from the output. These systems can't take corrective action to alleviate undesired changes in the output. Closed-loop systems (**Fig 1b**) are ones whose output is fed back and compared with the input in such a way as to maintain the desired output. This series of articles will consider only closed-loop systems.

You can represent any closed-loop system with the standardized diagram shown in **Fig 1b**. In the diagram, the output C (the directly controlled variable) feeds back through a functional block with transfer function H and is compared with reference signal R at a summing junction. The difference between R and the feedback signal (B) is the error, or actuating signal (E). The reaction of the components represented by G in response to error signal E maintains the output at the desired level.

If controlled variable C is fed back to the summing

An understanding of classical measurement methods for control systems aids you in using dynamic signal analyzers for design, modeling, analysis, and test.

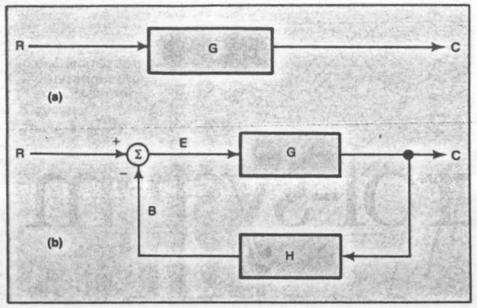

Fig 1—The two basic configurations for a system are the open-loop (a) and closed-loop (b) configurations. In an open-loop system, variations in the output go undetected; in a closed-loop configuration, feedback from the output maintains the desired output characteristics.

junction without any modification, transfer function H will equal one. A system in which H is 1—called a unity-feedback control system—is the easiest type of system for designers to analyze. These block diagrams will help you model a system. To evaluate the system's performance, however, you'll need criteria by which to judge it.

You need to know, for example, how quickly the system achieves the desired output level and whether the system can maintain that level with little or no variation. Control-system designers' need for this basic information led to the first serious study of negative-feedback control systems and ultimately to the development of classical linear control theory.

The first attempts to analyze systems, which were speed regulators on steam engines, took place in the 19th century and involved the use of differential equations. To make the equations as easy as possible to solve, engineers selected a small group of standard inputs that were easy to express mathematically. Chief among these inputs was the step forcing function. Not only was this function simple to express, but it was easy to implement physically, so it permitted the engineer to compare analytical results with measured results.

By using information derived from this step-response technique, you can measure both the speed with which a system reacts to a change in input (rise time) and the degree to which the system temporarily exceeds the desired output level (maximum overshoot). The information also indicates settling time, which is the time it takes for a system to reach a new output level within a given error band (Fig 2).

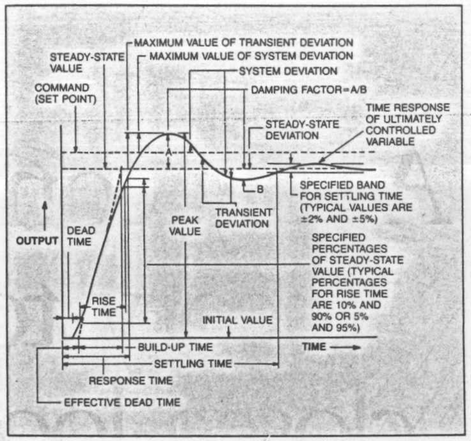

Fig 2—This output waveform, a typical response of a system to a step forcing function, can help you evaluate the stability of the system. The overshoot and ringing are good indicators of the system's stability. Settling time to within a given error band is an important parameter for linear amplifiers and D/A converters.

A system's settling-time spec indicates the relative stability of that system. For example, a system that oscillates around the steady-state value for long periods of time is considered to be less stable than a system whose oscillations die out relatively quickly. A system whose response oscillates indefinitely at either a constant or an increasing amplitude after the system has been excited by a step input is considered to be unstable (**Fig 3**).

Although measuring time-domain responses allowed engineers to verify system models and obtain useful, easily extracted performance information, this method provided few clues to how they could improve a system's performance. Further, as systems became more complex, the solution of the integrodifferential equations used to model them grew correspondingly difficult. Solving these equations became somewhat easier in 1899, however, when Oliver Heaviside introduced partial-fraction expansion techniques. Heaviside's discovery made it possible to use Laplace transforms to simplify the solution of large differential equations.

The Laplace transform

The Laplace transform is the integral of the product of the variables f(t) and e^{-st}. The variable s is complex; it has a real component (σ) and an imaginary component

($j\omega$). The transformation of a system's transfer function results in a ratio of polynomial expressions in s. In this format, many of the time domain's complex calculations become simple algebra problems.

Of particular interest to engineers are values of s that would cause either the ratio's numerator or denominator to equal zero. Values that set the numerator to zero force the function represented by the ratio to equal zero in the Laplace domain. These values of s are called zeros. Values of s that cause the denominator to equal zero force the function represented by the ratio to equal infinity in the Laplace domain. These values are called poles. Poles are of special significance: When you transform a function back into the time domain to predict the

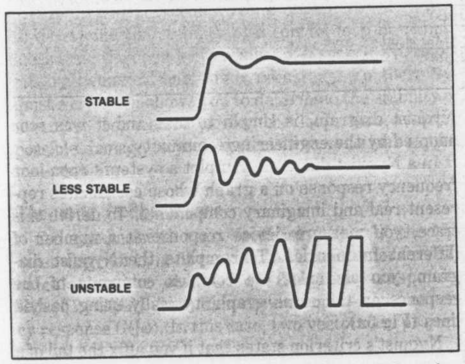

Fig 3—A system's settling characteristic in response to a step-function input shows how stable the system is. The more ringing there is around the waveform's final value, the less stable the system is. In an unstable system, the ringing never stops, but instead builds up to oscillation.

system's response, the real part of the pole determines the exponents of the response. If an imaginary part of the pole exists, it becomes a frequency component of at least one term of the response.

In addition, you can use the real part of the poles to determine the system's stability without having to transform the calculations back into the time domain. Poles with positive real values indicate positive exponents in the time domain; therefore, they indicate instability of the control loop.

These Laplace-transform methods were sufficient to solve control-system problems until the early part of the 20th century. With the development of the vacuum tube and large electronic systems, however, the computational aid of the Laplace transform became inade-

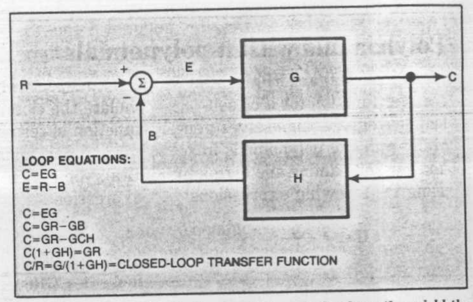

Fig 4—Simple multiplication, addition, and subtraction yield the transfer function for a closed-loop control system. The model applies equally well to locomotives and operational amplifiers.

quate for system evaluation. In particular, Harold S Black's use of negative feedback in electronic-amplifier design in the late 1920s spurred engineers to look for more powerful analytical tools.

From Laplace to Nyquist

In the early 1930s, Harry Nyquist discovered that the solution to the problem of analyzing complex systems lay in the frequency domain. His analysis was built upon the extension of several familiar concepts. By expressing the elements of the control loop as transfer functions (ie, the Laplace transform of the output divided by the Laplace transform of the input), simple algebra could provide an expression for the input/output relationship of a system.

By using the block diagram in Fig 1 and letting C, E, R, and B represent the Laplace transforms of the signals appearing at the lettered points in the system, you can express the input/output relationship of a closed-loop system as $C/R=G/(1+GH)$ (**Fig 4**). This expression is known as the closed-loop transfer function. Reduced to its simplest form, the expression is itself a ratio of polynomials in s.

From previous work in the Laplace domain, engineers already knew that the poles of this closed-loop transfer function would determine the stability of the system. If you look for the closed-loop poles (ie, values of s that force $1+GH$ to zero), you see that the GH term contains all the information regarding the poles' whereabouts (see **box**, "Polynomials within polynomials"). Therefore, the key to determining the stability of a system is knowing whether any of the values of s that make GH equal −1 (ie, the closed-loop poles) have positive real parts.

Closed-loop systems compare input signals with feedback signals in such a way as to maintain a precisely controlled output/input relationship.

Polynomials within polynomials

The terms G(s) and H(s) in the transfer function for a closed-loop system are themselves generally ratios of polynomials in s. You can thus define the terms by using the following expressions:

$$G(s) = \frac{G_n(s)}{G_d(s)}$$

$$H(s) = \frac{H_n(s)}{H_d(s)},$$

where the subscripts n and d indicate the numerator and denominator portions of G(s) and H(s), respectively. If you reformulate the closed-loop transfer function in terms of the numerator and denominator of G(s) and H(s), you obtain the following expression:

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + GH(s)} = \frac{\frac{G_n(s)}{G_d(s)}}{1 + \frac{G_n(s)\ H_n(s)}{G_d(s)\ H_d(s)}} = \frac{\frac{G_n(s)\ H_d(s)}{G_d(s)\ H_d(s)}}{\frac{G_d(s)\ H_d(s) + G_n(s)\ H_n(s)}{G_d(s)\ H_d(s)}}.$$

You express the closed-loop transfer function in this manner to illustrate that the term 1+GH(s) itself has poles and zeros, and that the zeros of this term determine the poles of the closed-loop transfer function. Note also that the zeros of the closed-loop transfer function are the roots of the equation $G_n(s)H_d(s)=0$.

If you could measure GH at each value of s you could simply determine when GH equals −1 and record those values of s. Unfortunately, the only values of GH you can physically measure are those values of s for which the real part equals 0—ie, GH(0+jω), which is commonly expressed as GH(jω). All other values of s simply provide an analytical model for understanding how a system's components will affect its time- and frequency-domain responses. Evaluating GH(jω) provides the input/output relationship of the device, typically in the form of gain and phase shift as a function of frequency. GH(jω), therefore, is called a frequency response, and it represents the frequency-domain characteristics of the system.

Nyquist's stability criterion

Using only the information from evaluating GH(jω), Nyquist had to determine whether there were any values of s (having positive real parts) that satisfied the equation GH(s)=−1. Of course, if GH(jω)=−1 at some frequency ω_1, then it's clear that a closed-loop pole exists at $s=(0+j\omega_1)$. But determining whether there were other closed-loop poles having positive real parts was not easy.

Nyquist's contribution lay in his discovery of a technique in which the closed-loop system's measured open-loop frequency response could be used to determine the existence of any closed-loop poles with positive real parts. The mathematical proof behind Nyquist's stability criterion is not intuitively obvious. However, the graphical representation of this discovery, known as a Nyquist diagram, is simple to use, and it was soon adopted by the engineering community.

In a Nyquist diagram, you plot a system's open-loop frequency response on a graph whose coordinates represent real and imaginary components. To derive this trace, you measure device response at a number of different frequencies. To complete the Nyquist diagram, you also plot the complex conjugate of the response on the same graph, typically using dashed lines (**Fig 5a**).

Nyquist's criterion states that if you affix the tail of a

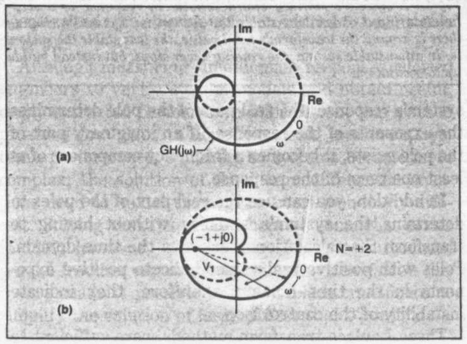

Fig 5—The Nyquist diagram (a), a tool for frequency-domain analysis of a system, *is a plot of the system's open-loop frequency response and its complex conjugate. To determine the number of closed-loop poles with positive real parts, count the number of net rotations of vector V_1 in* ***b.***

vector to the point representing −1+j0, allow the head of the vector to trace the entire path created by the open-loop frequency response and its complex conjugate, and then count the net revolutions (N) of that vector, you can determine whether the system has any closed-loop poles with positive real parts (**Fig 5b**).

The stability criterion actually determines the difference (N) between the number of zeros (Z) and the number of poles (P) of the term 1+GH(s), using the equation N=Z−P. In systems for which P=0 (which is often the case), N=Z, which is the number of poles that have positive real parts in the closed-loop transfer function.

The most important aspect of Nyquist's criterion is its usefulness. This tool provides engineers with a way to determine the stability of a control loop by simply measuring the open-loop frequency response. No rigorous mathematical analysis is necessary. Further, Nyquist's criterion allows you to determine the stability of a control loop before closing the loop, thus avoiding possible damage to your system from reactions caused by instability.

The unity-feedback case

For systems with unity feedback (H(s)=1), the Nyquist diagram (**Fig 6**) provides a technique for directly calculating the closed-loop frequency response, G(jω)/[1+G(jω)], from its measured open-loop frequency response G(jω). In this case, two vectors are projected—one from point −1+j0 (V_1) and the other from the origin (V_2)—to extend to the curve G(jω).

V_1 thus represents the term 1+G(jω), and V_2 represents G(jω). The vectors' heads are placed on the open-loop frequency response. You can use simple vector algebra (V_2/V_1) to calculate the input/output relationship of the closed-loop system at the corresponding frequency.

As an alternative to calculating the ratio of the vectors at many frequencies, graphical tools called magnitude contours and phase contours were devel-

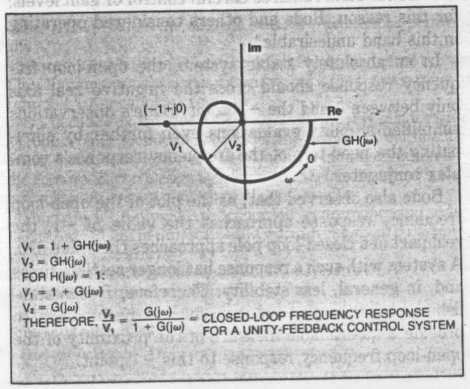

Fig 6—To evaluate unity-feedback closed-loop frequency response *from open-loop data, you use vector algebra with a Nyquist diagram. The ratio V_2/V_1 is the output/input response of the closed-loop system at any frequency.*

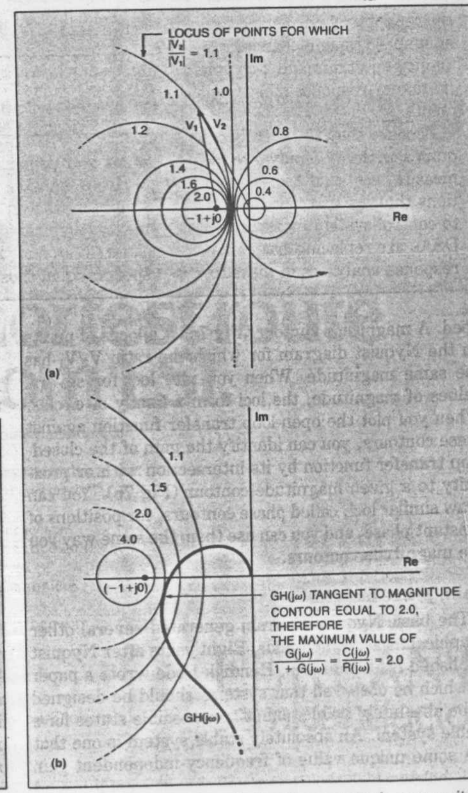

Fig 7—Loci of constant transfer-function magnitude, *or magnitude contours* ***(a)****, form a family of circles. By plotting an open-loop transfer function against the contours* ***(b)****, you can identify the gain of the closed-loop function by its intersection with the loci.*

Mathematical techniques by Laplace and Heaviside, and graphical-analysis methods by Nyquist, Bode, and Nichols aid you in predicting closed-loop system response.

Dynamic signal analyzers

Dynamic signal analyzers (DSAs) are low-frequency analyzers that sample signals applied to their inputs and then use Fourier transforms to analyze the signals in both the time and frequency domains. Useful to engineers who take mechanical, acoustic, and audio-electronics measurements, these analyzers record waveforms and take high-speed, single-channel frequency-spectrum and 2-channel frequency-response measurements.

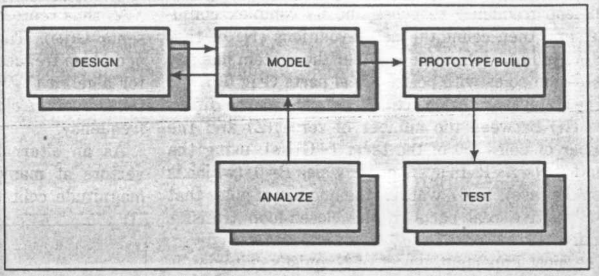

Fig A—In a typical product- or system-development cycle, these five processes all play crucial roles. A dynamic signal analyzer (DSA) expands the traditional role of test instruments from the analyzing and testing functions to include significant contributions to the model and design processes.

Because they've recently acquired an array of advanced measurement and analysis functions that are directly applicable to control-system measurements, DSAs are replacing frequency-response analyzers in measurement and performance analysis.

However, DSAs are no longer simply measurement instruments; they have new analytical capabilities that, in many respects, equal those of workstations (**Fig A**). These capabilities allow designers who use DSAs to measure a device under test and to model, revise, and perfect the response of the as-yet-unbuilt system.

oped. A magnitude contour (**Fig 7a**) is a locus of points on the Nyquist diagram for which the ratio V_2/V_1 has the same magnitude. When you plot loci for several values of magnitude, the loci form a family of circles. When you plot the open-loop transfer function against these contours, you can identify the gain of the closed-loop transfer function by its intersection with or proximity to a given magnitude contour (**Fig 7b**). You can draw similar loci, called phase contours, for positions of constant phase, and you can use them the same way you use magnitude contours.

Gain, phase margins

The basic Nyquist diagram generated several other graphical measurement aids. Eight years after Nyquist published his techniques, Hendrik Bode wrote a paper in which he observed that systems should be designed to be absolutely stable, one of two possible states for a stable system. An absolutely stable system is one that has some unique value of frequency-independent gain (K_1), below which the system will always be stable, and above which the system will always be unstable.

Alternatively, a conditionally stable system is one with a gain (K_2) above which the system will generally become unstable. In a conditionally stable system, there is also a band of gains greater than K_2 where the system becomes stable again. However, operating in this higher band requires careful control of gain levels; for this reason, Bode and others considered operating in this band undesirable.

In an absolutely stable system, the open-loop frequency response should cross the negative real axis only between 0 and the -1 point. Bode's observations simplified stability evaluations even further by eliminating the need to plot the frequency response's complex conjugate.

Bode also observed that, as the plot of the open-loop frequency response approaches the value of -1, the real part of a closed-loop pole approaches the value of 0. A system with such a response has longer settling times and, in general, less stability. Therefore, Bode established phase-margin and gain-margin parameters to provide a quantifiable measure of the proximity of the open-loop frequency response to this -1 point.

Bode viewed the Nyquist diagram as a polar plot in which gain is the distance from the origin and phase is measured as an angle with 0° as the positive real axis. The gain margin is the additional gain required to

Easy to implement in a test system, the step forcing function reveals a wealth of information about a system's speed, stability, and settling time.

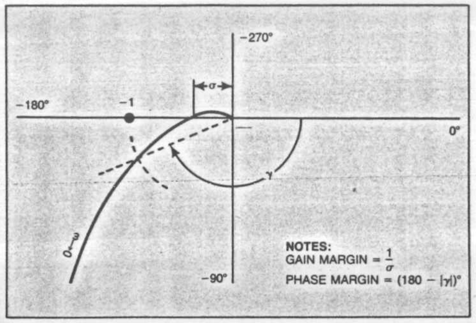

Fig 8—Measuring gain and phase margin with a Nyquist diagram is easy. The gain margin is the additional gain needed to achieve unity gain at 180° phase shift, and the phase margin is the additional phase shift needed to achieve 180° at the unity-gain frequency.

achieve unity gain at the frequency where the phase of the open-loop frequency response equals 180° as it crosses the negative real axis (**Fig 8**). Simply stated, the gain margin is the additional gain required to place a closed-loop pole on the jω axis (GH(jω)=−1) and produce instability.

Gain margin alone is not sufficient for you to determine how close GH(jω) is to the −1 point and, therefore, how close a closed-loop pole is to the jω axis, as **Fig 8** shows. You also need to use phase margin, which is the additional phase shift required to achieve 180° of phase shift at the highest frequency at which the open-loop gain is unity. Typical target values are a gain margin of not less than two (6 dB) and a phase margin of not less than 30°.

The Nyquist diagram is a powerful tool for analyzing all types of systems, but it is not without shortcomings. Its primary limitation is that it lacks a convenient graphical technique that you can use to predict how changes in the system will affect the open-loop frequency response. Any time the system is changed, you must remeasure the response or recalculate it from your system's model. Bode developed a diagram that alleviates this problem.

The Bode diagram (**Fig 9**) is also a plot of the open-loop frequency response, except that Bode treated gain and phase separately, plotting each as a function of frequency (ω). This technique has several advantages for the designer who must evaluate systems while developing them.

By expressing gain in logarithmic units (dB) and phase in degrees, the Bode diagram allows you to

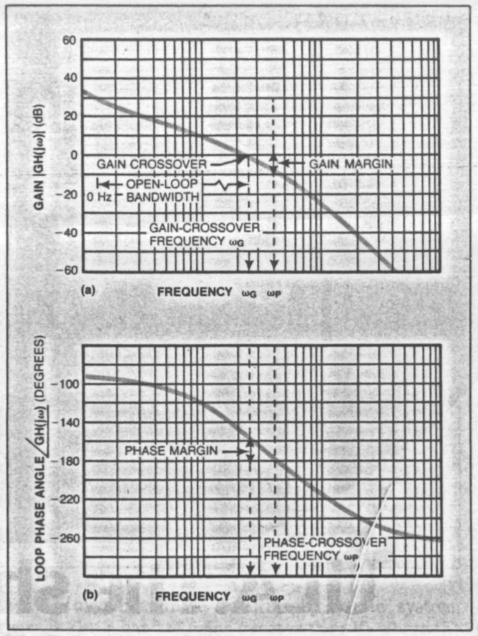

Fig 9—A Bode plot yields gain and phase margin and bandwidth. The open-loop bandwidth is defined here as the frequency at which the system has unity gain.

combine new data with already established measurements through simple addition. The resulting trace is equivalent to the frequency response you would obtain by connecting the actual devices in cascade and physically measuring the composite.

Because of certain techniques that Bode developed, engineers could sketch an approximation of a device's frequency response from its transfer function—ie, they could evaluate GH(jω) over a range of frequencies from the equation for GH(s)—in a Bode diagram without taking the physical measurements. Designers could also perform the opposite function, that of approximating a transfer function from a measured response. For the first time, engineers could link a system's model to measured data without having to resort to extremely laborious calculations.

Bode plots like the ones in **Fig 9** let you measure a system's gain and phase margins. A third performance parameter, the open-loop bandwidth, is also easy to

Graphical-analysis techniques allow you to use a negative-feedback system's open-loop response to predict the system's closed-loop response.

measure on a Bode plot. Open-loop bandwidth indicates how fast a system can react to a change in input; this measurement is analogous to rise time in the time domain. Specifically, open-loop bandwidth is the span between 0 Hz and the frequency at which the open-loop response has unity gain (0 dB). The greater the open-loop bandwidth is, the faster the system will react.

No graphical technique using a Bode plot exists for directly calculating the closed-loop frequency response of unity-feedback systems. Although such a technique (the use of magnitude and phase contours) does exist for the Nyquist diagram, this diagram's linear scales would seldom accommodate both the range of gain values produced by a system and the level of detail about the unity-gain point required to determine stability.*

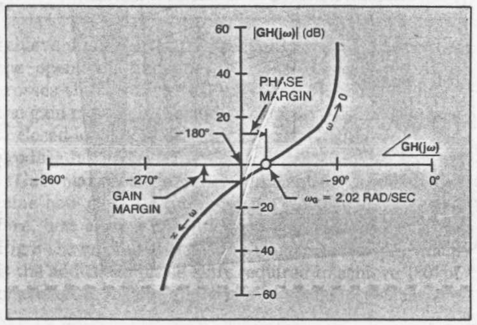

Fig 10—By plotting logarithmic gain against linear phase in a Nichols diagram, you can calculate graphically a unity-gain system's closed-loop frequency response from the open-loop response.

The Nichols diagram, a third graphical analysis tool, was developed specifically to accommodate the graphical calculation of the closed-loop frequency response of a unity-feedback system from the system's measured open-loop frequency response. The Nichols diagram uses both the single-plot concept of the Nyquist diagram, in which frequency is a varying parameter (but not a coordinate), and the logarithmic gain and linear phase scales of the Bode plot. Thus, the Nichols diagram is a plot of logarithmic gain versus linear phase as a function of frequency, as **Fig 10** shows.

Because of the wide range of gains represented by the logarithmic scale, however, the open-loop frequency response of almost any system could be plotted on a standardized grid. It was practical, therefore, to superimpose a second grid (of magnitude and phase contours)

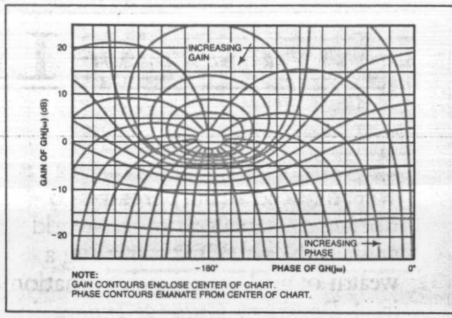

Fig 11—Magnitude and phase contours are superimposed on a Nichols diagram, forming a Nichols chart. The chart allows you to make a quick estimate of the closed-loop response of a unity-feedback system.

on the Nichols diagram for calculating the closed-loop frequency response. The resulting Nichols chart (**Fig 11**) rapidly became a standard graphical tool for quickly estimating the closed-loop frequency response of a unity-feedback system. Unfortunately, like the Nyquist diagram, the Nichols chart doesn't make it easy for you to combine frequency responses, nor does it provide graphical tools for linking system models and frequency responses. **EDN**

Authors' biographies

Steven W S Asbjornsen is a product-marketing engineer at Hewlett-Packard (Everett, WA), where he's responsible for control-system application and market development. He received a BSEE from the University of Washington and has been with HP for five years. Steve, who is a musician in his spare time, also enjoys hiking and waterskiing.

Owen Brown, who is in charge of product publicity for Hewlett-Packard (Palo Alto, CA), has worked for HP for two years. He has a bachelor's degree from Yale University and an MBA from the University of Chicago. His spare-time pursuits include painting and aikido.

Article Interest Quotient (Circle One)
High 482 Medium 483 Low 484

PLACE
STAMP
HERE

Texas Instruments Incorporated
P.O. Box 809066
Dallas, Texas 75380-9066

PLACE
STAMP
HERE

Texas Instruments Incorporated
P.O. Box 809066
Dallas, Texas 75380-9066

PLACE
STAMP
HERE

Texas Instruments Incorporated
P.O. Box 809066
Dallas, Texas 75380-9066

Dynamic signal analyzers simplify measurement of linear control systems

Advanced dynamic signal analyzers (DSAs) give designers a wide choice of techniques for measuring a system's open-loop frequency response. This article, part 2 of a 3-part series, considers the effect of DSAs on the graphical measurement techniques of linear control-system theory. Part 1 of the series presented an overview of classical linear control theory. Part 3 will explore the expanded role of DSAs in the control-system design process.

Steve Asbjornsen and Owen Brown,
Hewlett-Packard Co

You can use a variety of graphical techniques to analyze a negative-feedback, closed-loop control system (**Ref** 1). Bode plots, Nichols diagrams, and Nyquist diagrams are three distinct tools that allow you to determine a system's amplitude and phase characteristics (and, therefore, its stability) as functions of frequency. Typical test instruments make use of these tools, and a dynamic signal analyzer (DSA) is no exception—it allows you to view and plot frequency-response data in all three formats.

The internal analysis functions of modern DSAs, however, alter the relative usefulness of these three graphical techniques. Understanding these DSA functions gives you an idea of how a DSA's computational power can expand your test options.

Waveform math and Nyquist diagrams

A DSA's waveform-math capability, for example, limits the Nyquist diagram's usefulness to providing a complete check of a system's stability and a 1-trace representation of its frequency response. The waveform-math utility is a built-in calculator that allows you to add, subtract, multiply, divide, or use any of the other operators shown in **Table 1** to manipulate frequency responses, recorded waveforms, and complex constants.

The Nyquist diagram lets you easily determine the stability of all types of systems, including absolutely and conditionally stable systems. You can also directly calculate the closed-loop frequency response of a unity-feedback control system. However, the diagram doesn't facilitate calculation of composite frequency responses, and its linear scales can't accommodate both adequate gain ranges and acceptable resolution around the unity-gain point. Reading phase margin is more difficult with the Nyquist diagram than with other diagrams, and reading the open-loop bandwidth is

A DSA's built-in calculator uses waveform math to perform arithmetic operations and to manipulate frequency responses, waveforms, and complex constants.

TABLE 1—WAVEFORM-MATH FUNCTIONS IN A DSA

ADD	SQUARE ROOT	MULTIPLY BY jω	T/(1−T)
SUBTRACT	RECIPROCAL	FFT	REAL PART
MULTIPLY	NEGATE	INVERSE FFT	COMPLEX CONJUGATE
DIVIDE	DIFFERENTIATE		LOG DATA

impossible unless the frequency at which the gain becomes unity is recorded on the plot. Finally, you can't use the Nyquist diagram to estimate the transfer function of a system from its measured frequency response, or vice versa.

A DSA's waveform-math utility, however, lets you calculate a system's closed-loop response precisely, in any format, using the equation

$$C(j\omega)/R(j\omega)=G(j\omega)/\{1+G(j\omega)\}.$$

Using waveform math, you can calculate closed-loop frequency response independently of the display format. Linear scales are not a problem when you're using a DSA, because the DSA's display can easily rescale the data. The DSA's marker readouts make the measurement of gain margin, phase margin, and open-loop bandwidth much easier.

Bode plots

DSAs also alter the usefulness of the Bode plot, which designers have traditionally favored because they could use it to estimate composite frequency responses quickly. The Bode plot's logarithmic units offer a large dynamic range of gains, and the plot makes it easy to measure gain margin, phase margin, and open-loop bandwidth. Finally, unlike the Nyquist diagram, the Bode diagram lets you estimate a transfer function from a frequency response and vice-versa.

The Bode plot's major drawbacks are that you have to plot traces for both gain and phase, and that you can't estimate the closed-loop frequency response from the open-loop frequency response. A DSA's waveform-math utility makes it easy to calculate the closed-loop frequency response, however. Further, the DSA's frequency-response-synthesis and curve-fitting functions automate the transition between frequency responses and transfer functions.

The Bode plot is still useful in that it helps you intuitively understand the frequency-response/transfer-function transition. The Bode plot also helps you estimate composite waveforms, and its logarithmic gain units provide both range and resolution.

Although the Bode plot and Nyquist diagram are still useful to designers who perform system analysis on DSAs, Nichols diagrams are not. The Nichols diagram's only advantage over the other diagrams is that it lets you calculate closed-loop frequency responses. Because the waveform-math capability of DSAs solves this problem, it renders the Nichols diagram obsolete.

The root-locus plot

The newest method of making control-system measurements is root-locus analysis, which was developed in the late 1940s and early 1950s by Walter R Evans. All previous methods of analysis had used open-loop frequency response solely to determine whether closed-loop poles with positive real parts existed. These methods yielded no additional information concerning the actual value of s for the poles. The root-locus technique, however, lets you examine the actual values of s for the closed-loop poles graphically, based on the known values of s for the open-loop poles and zeros.

The root-locus diagram could not have been conceived without the development (in the late 1940s) of the s-plane. The s-plane is a 2-dimensional plane that represents all possible values of the Laplace variable s. The plane's ordinate is the imaginary part (ω of s=σ+jω), and its abscissa is the real part (σ of s=σ+jω), of s (**Fig 1**).

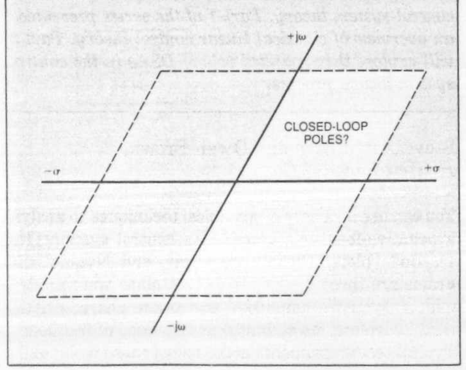

Fig 1—The s-plane represents all possible values of s in two dimensions. This plot allows you to locate the poles and zeros of a closed-loop transfer function and thereby determine the stability of the system characterized by the transfer function.

Each value of s, therefore, has a unique position in the s-plane. If you can determine the poles and zeros of a ratio of polynomials (such as the open- and closed-loop transfer functions of a control system) in s, you can plot the location of these poles and zeros in the s-plane. Any closed-loop pole that exists in the right half of the s-plane represents poles with positive real parts and therefore indicates an unstable system.

In measuring the open-loop frequency response of a system, you're collecting the same data you would if you were evaluating GH(s) for values of s that lie on the positive ordinate of the s-plane (s=0+jω for ω=0 to +infinity). From the open-loop information, Nyquist—without using the idea of the s-plane—made the conceptual leap that allowed him to determine whether there were any values of s with positive real parts that solved the equation GH(s)=−1. His observation was not an obvious one, to say the least.

Consider a control loop that has been opened so that the open-loop frequency response (GH(jω)) is measurable. If you alter just the gain of the loop, you won't affect the value of the open-loop poles and zeros. As a result, the open-loop transfer function can be expressed as GH(s)=KGH(s), where K represents a proportional gain constant that's independent of s.

Although varying K has no effect on the position of the open-loop poles and zeros, it can have a tremendous effect on the closed-loop poles. This effect becomes apparent if you substitute KGH(s) in the denominator of the closed-loop transfer function G(s)/{1+GH(s)} and solve the equation for the poles. The resulting expression is 1+KGH(s)=0.

The closed-loop poles, therefore, are values of s that are solutions to the equation GH(s)=−1/K. To locate the closed-loop poles, you must, whenever K changes, find new values of s that satisfy the equation GH(s)=−1/K. It was this relationship—between the stationary poles and zeros of the open-loop transfer function, and the closed-loop poles that vary with pure gain—that provided the basis for Evans's root-locus technique.

Using the root locus

The root-locus technique plots the open-loop poles and zeros in the s-plane. You can obtain the open-loop pole and zero locations from a mathematical derivation of the open-loop transfer function or by using Bode's techniques to extract the transfer function from a measured frequency response. If you plot the open-loop poles and zeros on the s-plane (**Figs 2a and 2b**), you can

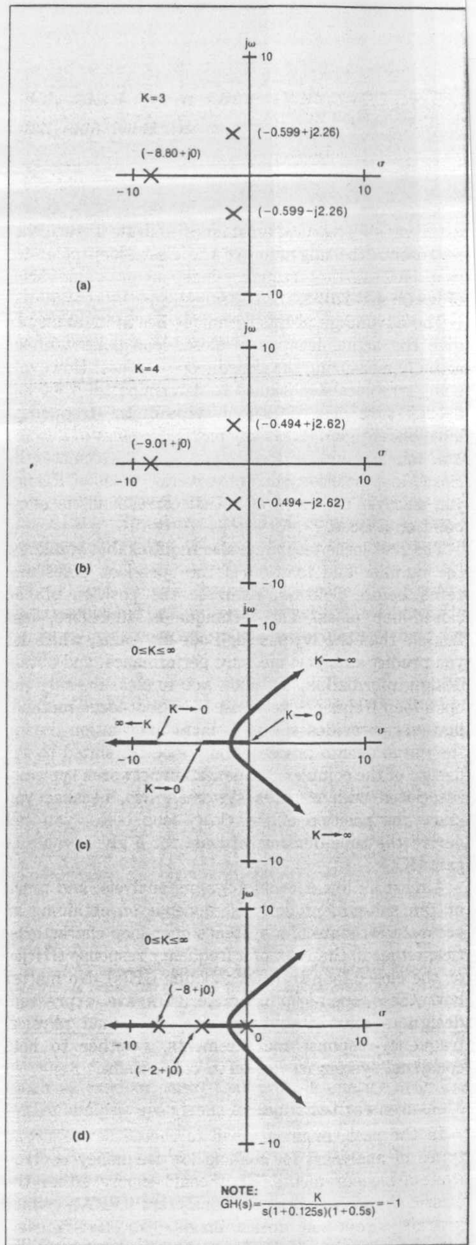

Fig 2—A powerful graphical technique, the root-locus plot, allows you to determine the location of poles in a closed-loop system without actually measuring the closed-loop response. This method depicts the migration of poles as the system's frequency-independent gain (K) varies.

use Evans's graphical techniques to draw a trace that represents the migration of the closed-loop poles (or root loci) as the frequency-independent gain varies (**Figs 2c** and **2d**).

The advantage of this technique lies in its ability to give the actual location of closed-loop poles without actually measuring the closed-loop response. However, using graphical techniques to determine the root loci doesn't give you the actual value of the frequency-independent gain (K) at any particular point in a locus. You must return to the equations and calculate the closed-loop pole locations for several values of K until you discover the value of K that corresponds to some point on a locus.

The root-locus technique also requires that you know the number and location of the open-loop poles and zeros before you can estimate the position of the closed-loop poles. The technique is, therefore, less flexible than the Nyquist or Bode diagrams, which let you predict stability, measure performance, and obtain design information, but allow you to measure only the open-loop frequency response. The root-locus method, however, provides you with more information during the initial design process, and it's better suited to the design of the complex compensation networks typically associated with complex systems. Also, because you know the position of the closed-loop poles, you can derive the time-domain response for a given value of gain (K).

Almost all linear control-system analysis, and much of the subsequent designs, depends on obtaining an accurate estimate of a system's open-loop characteristics, either in the form of a frequency response $GH(j\omega)$ or the closed-loop transfer function $GH(s)$. No matter how these open-loop characteristics are expressed, designers must always perform the actual physical frequency-response measurements, whether to help construct system models or to verify them.

Measurement techniques

In the past, engineers had to choose between two types of analyzers for making low-frequency control-system measurements. They could choose either the classic frequency-response analyzers (FRA), which provide swept-sine measurements, or fast-Fourier-transform analyzers (FFTA), which can measure a whole spectrum in one measurement.

The two analyzers have different advantages and disadvantages. For example, although the FFTA has the potential for faster measurement times, it entails complex set-up procedures. And although the FRA is familiar to engineers, who understand its swept-sine method of measurement, its measurement times are slow. However, because DSAs offer both measurement techniques, designers no longer have to accept the tradeoffs that accompany choosing an FFTA or an FRA.

Frequency-response analyzers

FRAs operate in much the same way as do heterodyne network analyzers, and they're limited to taking measurements at low frequencies. They generally possess two channels, each of which uses a discrete Fourier transform to emulate a single bandpass filter. The Fourier integration time controls the filter's bandwidth to values in the low microhertz range, and an integrated sine-wave source (**Fig 3**) synchronizes the filter's center frequency.

A stimulus signal from the FRA drives the device under test. The analyzer's two channels connect to the input and output of the device, and the signal each channel receives undergoes comparison with the stimulus signal as a function of the discrete Fourier transform. The result is a complex value containing the magnitude and phase (with reference to the stimulus signal) of the measured signal.

The FRA then compares the two channels' results, deriving the gain and phase-shift relationships between the two channels' signals. This process occurs several

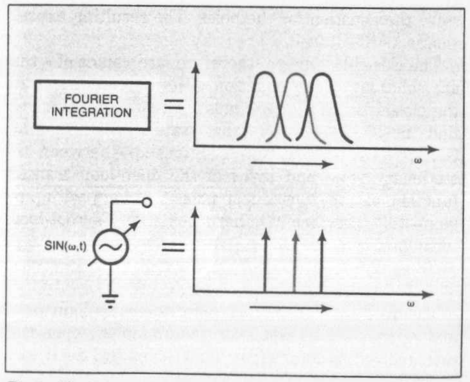

*Fig 3—**The integration time of a Fourier transform** controls filter bandwidth, and a sine-wave signal source sets the filter's center frequency. Each discrete Fourier transform emulates a single bandpass filter.*

A root-locus diagram allows you to examine the values of s for a system's closed-loop poles, based on the known values of s for the open-loop poles and zeros.

times between the start and stop frequencies being analyzed, thereby producing a series of discrete gain and phase values. When you connect these points graphically, you obtain the gain and phase curves of the frequency response. Note that this measurement process has two implementations, and the difference between them can be important to the designer who uses computer-aided analysis.

The primary distinction between the two implementations lies in the sources. An FRA whose source sweeps continuously can integrate the signal while the source is actually sweeping. Each integration period, therefore, covers a small part of the total measurement span. The result of that integration is then available at the end of each integration period.

This continuous-sweep technique creates a potential ambiguity between the phase and magnitude values for the displayed frequency, and the exact frequency at which they occurred. The ambiguity becomes especially serious when the integration period covers large frequency spans. Because the integration period is typically fixed, you can generally minimize this problem by reducing the sweep speed—and therefore the frequency span covered—during integration. The ambiguity won't interfere with graphical analysis, but it can create difficulties in computer analysis.

Sweep-and-dwell sources

The alternative to the continuous-sweep implementation is a sweep-and-dwell sine-wave source. In this type of analyzer, the sine-wave source dwells at a discrete frequency during the integration process and then performs a phase-continuous sweep to the next analysis frequency.

Because a sweep-and-dwell analysis occurs at a discrete frequency, the phase and gain analyses apply only for the frequency point at which the measurement was made; therefore, no ambiguity exists. DSAs incorporate this sweep-and-dwell form of swept-sine analysis, which optimizes the accuracy of their integrated computer-aided-analysis functions.

One possible drawback to the sweep-and-dwell technique is that the analyzer might miss valuable information between measurement points. However, by simply decreasing the sweep rate of these analyzers, you increase the number of measurement points between the start and stop frequencies and provide better resolution.

Newer analyzers offer an autoresolution function that monitors the gain and phase shift between measurement points and automatically adjusts the resolution (ie, sweep rate) during the sweep, thereby preventing the loss of valuable data. This function can also minimize total sweep time by increasing the sweep rate in portions of the frequency response that are relatively flat in both gain and phase.

FFT analyzers

Fast-Fourier-transform analyzers (FFTAs) are similar to FRAs in that they use a type of Fourier transform to achieve narrow analysis bandwidths. Their method of signal generation and use of two channels to compare a device's input and output are also the same as those of the FRAs. However, instead of emulating a single bandpass filter and tracking it over the spectrum of interest, FFTAs emulate hundreds of bandpass filters (**Fig 4**) and provide complete coverage of an entire spectrum in one integration period. FFTAs can usually perform measurements much more quickly than can FRAs.

In addition to the increased number of analysis bands, the FFT process can also use a wide range of

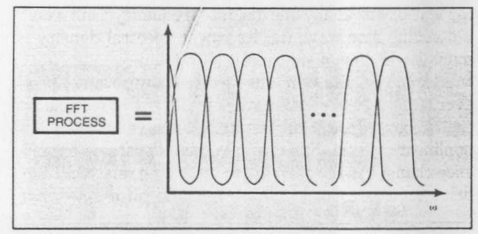

*Fig 4—**FFT analyzers emulate hundreds of bandpass filters.** One integration period, therefore, provides complete coverage of an entire spectrum. The FFTA can use a wide variety of stimulus signals, including random noise.*

stimulus signals. They typically use stimulus signals (such as random noise) that provide energy over the entire analysis span, thus taking full advantage of the analysis power.

FFTAs and FRAs use different methods to reduce measurement noise. If you don't know what the differences in the methods are and how they affect the measurement process, you can very easily misuse an FFT analyzer. Designers lacking this information have sometimes concluded—mistakenly—that FFT analyzers can't make control-system measurements.

Although the term "swept-sine analysis" describes the FRA's stimulus signal, it doesn't describe the

The sweep-and-dwell sine-wave stimulus is better than any other type of stimulus for measuring noisy systems.

FRA's unique analysis process comprehensively (heterodyne analyzers also use a swept-sine stimulus). The term "swept Fourier analysis" (SFA) describes the FRA's measurement process more specifically. The differences between FFT and SFA measurement processes lie mainly in their stimulus signal, single- versus multiple-band analysis, and noise-reduction techniques.

Stimulus signals in SFAs and FFTs

The SFA measurement process uses a swept-sine-wave stimulus, and the FFT process uses stimuli that produce energy at all the analysis frequencies within a single integration period. When you're measuring extremely noisy systems, the type of stimulus itself can have a profound effect upon the measurement.

The sweep-and-dwell sine-wave stimulus is better than any other type of stimulus for measuring noisy systems, because the power of the stimulus is concentrated at one discrete frequency. This concentrated-power approach automatically provides the best possible signal-to-noise ratio without any signal processing. A random-noise stimulus, on the other hand, must distribute its energy over a wider bandwidth, providing less power at any one discrete frequency than would a dwelling sine wave (ie, its power spectral density is much lower than a sinusoid's) (**Fig 5**).

A random-noise stimulus also has advantages, however. One of the key strengths of this type of stimulus is that it provides a linear estimate of the operation of a nonlinear system. For example, many systems experience changes in their frequency response relative to the drive level or relative to the direction of a sine-wave sweep. Random noise, which has no sweep direction and has random amplitudes at all frequency components, provides an average of the drive-level and sweep-direction effects, so it usually provides a good approximation of a system's operation.

In situations in which initial measurements indicate that the energy level of the random-noise stimulus is too low, you can improve the relative power spectral density of the stimulus by reducing the frequency span of the measurement (if the analyzer uses a band-limited random-noise source). However, to cover the original frequency span of interest, you must take more measurements.

Single- vs multiple-band analysis

The SFA's single-filter measurement process is slower than the FFT process, which provides hundreds of filters. However, the use of a single filter does have its advantages. If you use a single filter, you can make all the signals produced by distortion products (such as harmonic distortion and intermodulation distortion) lie outside the analysis bandwidth of a single filter, thus removing the products from the measurement.

By increasing the Fourier integration time, you can always reduce the filter's bandwidth to exclude distortion products (**Fig 6**). The only time you can't remove a disturbance signal is when a spur at a fixed frequency occurs at exactly the same frequency as that of the SFA's stimulus. Because of its many filters, the FFT process can be affected by distortion products, depending on the stimulus used.

For example, if you use a sine-wave stimulus in a

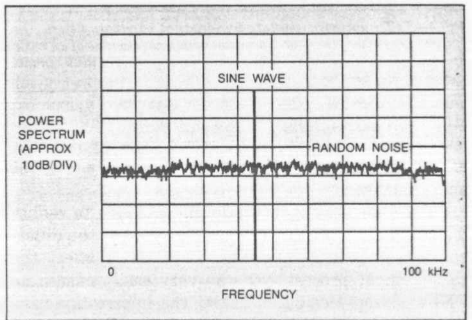

***Fig 5—Power-spectrum density for a random-noise signal** is much lower than that for a sine wave, as this plot shows. The random-noise signal distributes its energy over a much wider bandwidth than does a sine wave.*

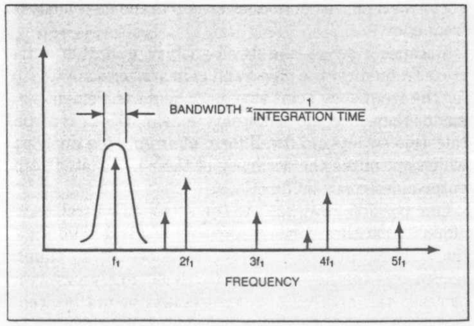

***Fig 6—Reducing the analysis band of an SFA's measurement process** excludes distortion products from the measurement. The analysis band is inversely proportional to the Fourier integration time.*

A DSA combines the advantages of classic frequency-response (swept-sine) analyzers and fast-Fourier-transform analyzers.

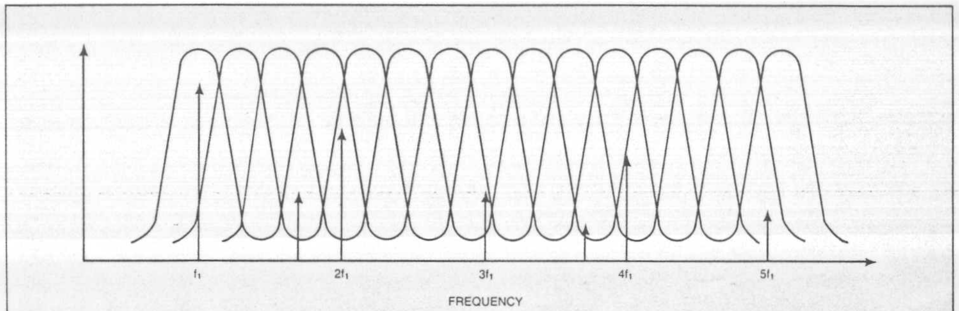

Fig 7—A sine-wave stimulus in a measurement using the FFT process can cause distortion products to appear within the filters produced by the transforms. The products would thus be recorded as part of the system's response.

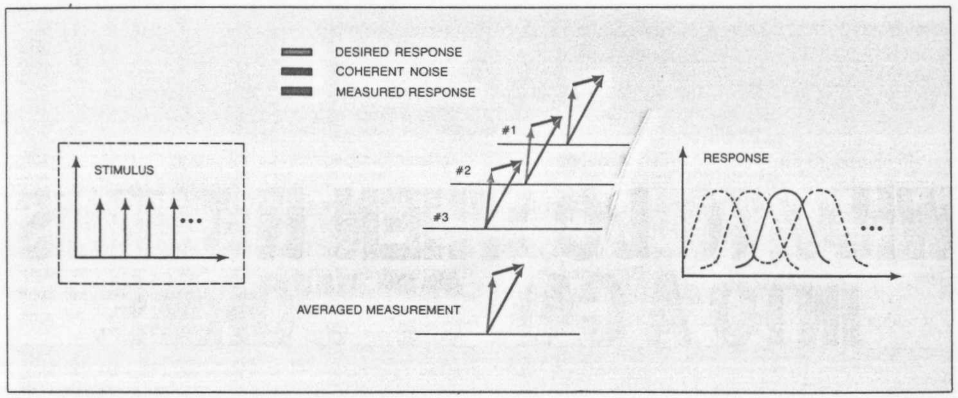

Fig 8—Periodic stimuli allow harmonics to maintain a constant phase relationship with desired signals. It's impossible, therefore, to average these components to zero, thereby reducing the effect of the nonlinearity on a measurement. A random stimulus eliminates this distortion-induced problem; averaging causes the nonlinear portion of the response in each filter to dwindle to zero.

system that produces harmonic distortion, the distortion products can appear within one or more of the FFT's filters and be recorded as part of the system's response (**Fig 7**).

Distortion products can also affect the FFT process when you use a pseudorandom signal as a stimulus. You can characterize this pseudorandom signal as a summation of discrete sine waves, each of which is tuned to the center frequency of a unique filter. If you separate the response in each filter into the portion of the response that results from the intended stimulus (the desired response) and the portion of the response that arises from a harmonic product of a lower frequency, you'll never see a change in the relationship between the desired response and the distortion product from measurement to measurement. Therefore, even if you were to average the results of several measurements, you wouldn't reduce the effect of the nonlinearity on the measurement (**Fig 8**).

A random stimulus, however, eliminates this distortion-induced problem by letting the nonlinear portion of the response in each filter decay to zero with averaging, even if the nonlinearity is a fixed spur at the center frequency of a filter. A swept-sine-wave stimulus

An FFT analyzer uses two channels to compare input and output and emulates hundreds of bandpass filters to provide complete coverage of an entire spectrum.

doesn't permit such averaging. The explanation of distortion phenomena leads directly to the topic of noise reduction.

Noise reduction in SFAs and FFTs

The noise-reduction process of SFAs is fairly straightforward. If you increase the integration time in each channel, the analysis bandwidth in each channel becomes smaller and smaller while remaining centered on the frequency of the stimulus. As the bandwidth becomes smaller, the noise power (of the measured system) within the filter lessens. Moreover, any spurs close to the center frequency become located farther down the stop band of the analysis bandwidth until they receive sufficient rejection.

The only type of distortion that can't be rejected is a spur that has the same frequency as the stimulus and that maintains a constant phase relationship with the stimulus. In this case, the spur is said to be coherent with the stimulus. A key aspect of this type of noise reduction is that the noise in each channel is reduced before the gain and phase relationship (ie, the frequency response) between channels is calculated.

The FFT process makes two types of noise reduction available: time averaging (a form of linear averaging) and power-spectrum averaging. Time averaging is very similar to the SFA's averaging process in that it improves the S/N ratio of the signal in each channel before the frequency response is calculated. The time-averaging method gathers the samples of the signal normally considered by the FFT into blocks of data called time records, and then averages the time records.

To keep averaging from reducing the signal of interest, you must make sure that the signal is a periodic one (such as the pseudorandom signal mentioned above) and that the phase of the signal is the same in each time record. You must also supply a trigger signal to the analyzer to indicate when data collection should begin.

Power-spectrum averaging

The second, and more commonly used, form of noise reduction in FFT measurements is power-spectrum averaging. The fundamental difference between this technique and time averaging is that relatively little

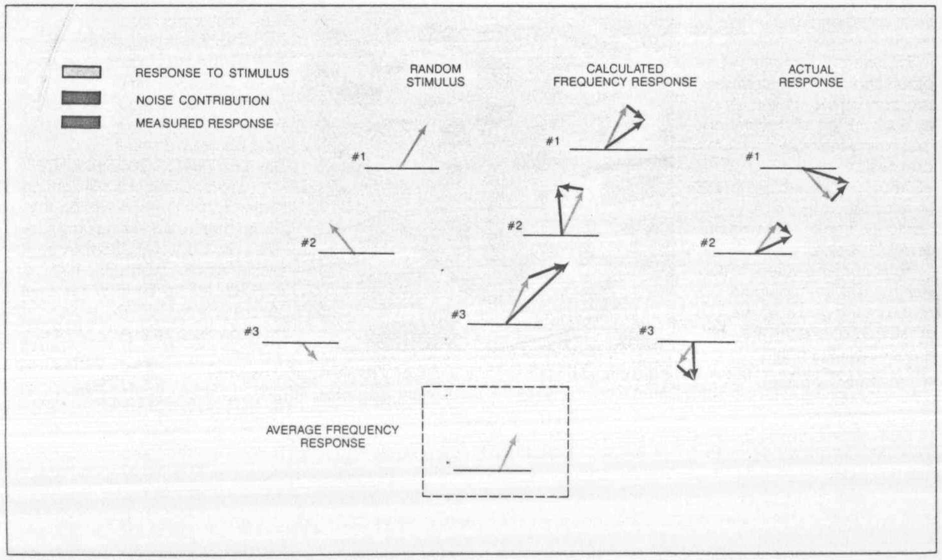

*Fig 9—**If you average frequency-response calculations,** you can use random noise as a stimulus. The stimulus-response relationship remains constant, while the noise contribution in both channels averages to zero. The averaged frequency response thus converges in a linear representation of the system's frequency response.*

If you don't understand the differences in the ways an FFTA and an FRA effect measurement-noise reduction, you could misuse the FFTA.

noise reduction occurs in each channel. Instead, the method removes noise by averaging frequency-response data from each measurement.

The benefit of this noise-reduction technique is that it doesn't matter how much the stimulus signal differs from one measurement to the next, as long as the gain and phase relationship from measurement to measurement remains the same. You can thus use random noise in an averaged measurement.

For example, assume you allow the random-noise stimulus to maintain the same frequency components as the pseudorandom-noise signal discussed earlier, but let the phase of the discrete sine waves vary randomly (a poor representation of random noise, but useful for this example). In this case, the stimulus signal within each filter will have a different phase orientation in each measurement.

If you were to examine a filter whose output is composed of a linear response to the intended stimulus signal and a nonlinearity-induced harmonic product, you'd see a change from measurement to measurement in the relationship between the linear response and the harmonic. The disparity exists because the phase of the harmonic's fundamental and the intended stimulus would have changed between measurements.

If, over several measurements, you examine a vector representing the computed frequency-response data for each measurement, you'll see that the distortion product appears as a vector that rotates about the end of a stationary frequency-response vector **(Fig 9)**. When you average several frequency-response vectors, the contribution from the distortion product falls to zero. The averaged frequency-response vector thus gives you the best linear estimate of the device's frequency response. If you use a random-noise stimulus in this type of averaging scheme, you'll find that even a spur at the center frequency of a filter would be noncoherent with the stimulus and would average to zero.

Although power-spectrum averaging, combined with a random-noise stimulus, reduces the effects of all forms of distortion products from a measurement, you wouldn't benefit from using power-spectrum averaging with a periodic stimulus. Using a periodic stimulus would allow distortion products to be coherent with the stimulus, so they'd be unaffected by averaging. Further, certain control-system measurements don't allow the use of power-spectrum measurement.

The FFT analyzer is always better than a swept-frequency analyzer for measuring a basically linear system with poor to good S/N conditions. Both analyzers (FFTA and SFA) will provide the same response, but the FFT process will provide it much more quickly. The SFA, on the other hand, gives you the best possible S/N ratio, so it's more suitable for use in difficult measurement situations. Having both techniques available is clearly preferable, as in a DSA, so you can use them to handle different measurement problems. **EDN**

Reference

1. Asbjornsen, Steve, and Brown, Owen, "Apply control-system theory to analyze closed-loop system," *EDN*, April 3, 1986, pg 173.

Authors' biographies

Steven W S Asbjornsen is a product-marketing engineer at Hewlett-Packard (Everett, WA), where he's responsible for control-system application and market development. He received a BSEE from the University of Washington and has been with HP for five years. Steven, who is a musician in his spare time, also enjoys hiking and waterskiing.

Owen Brown, who is in charge of product publicity for Hewlett-Packard (Palo Alto, CA), has worked for H-P for two years. He has a bachelor's degree from Yale University and an MBA from the University of Chicago. His spare-time pursuits include painting and aikido.

Article Interest Quotient (Circle One)
High 482 Medium 483 Low 484

Analyzers aid in the design of closed-loop systems

You can use the various functions of a dynamic signal analyzer to derive the open-loop frequency response of a system from measurements taken at various points in the system. This article, Part 3 of a 3-part series, discusses how to use DSAs to develop and model a control system. Part 1 of the series presented an overview of classical linear control theory, and part 2 considered the role of DSAs in control-system measurement.

Steve Asbjornsen and Owen Brown,
Hewlett-Packard Co

By manipulating the classical graphical-analysis tools of linear control theory, a dynamic signal analyzer (DSA) can greatly assist you in designing and modeling stable closed-loop, negative-feedback control systems. You can use a DSA's internal waveform-math functions to calculate the frequency response of a closed-loop control system by measuring its open-loop frequency response (**Refs 1** and **2**). By analyzing a closed-loop system's open-loop response, you can tell whether the system is stable or unstable, and you can design compensation networks to stabilize an inherently unstable system.

DSAs allow you to use three separate techniques for measuring the open-loop frequency response of a closed-loop system: the loop-open direct method, the loop-closed direct method, and the loop-closed calculated method. In the loop-open direct method (**Fig 1**), which is defined by the expression $B(j\omega)/E(j\omega)$, you open the loop by removing the summing junction from the system. You then inject a stimulus signal at point $E(j\omega)$ and measure the response at point $B(j\omega)$.

The ratio $B(j\omega)/E(j\omega)$ provides the gain and phase characteristics of $GH(j\omega)$, the open-loop frequency response. This technique lets you determine the stability of the loop before you close it. Before taking this measurement, however, you must take steps to avoid four problems: overdriving the system, changing loading conditions, system saturation, and loss of operation.

Overdriving the system

When you stimulate a high-gain system, you must be careful not to overdrive the system, that is, to exceed any part of the system's maximum operating range (a situation that might arise, for example, when a sinewave stimulus sweeps through a resonance). If you exceed this range, you could introduce nonlinearities into the measurement. Furthermore, if the stimulus contains high-energy components, you could damage the system.

Fortunately, you can often avoid overdrive by using a DSA's monitoring functions. The monitoring functions provide source-level control that can vary the stimulus level to maintain a specified input level to either channel of the analyzer.

You must pay attention to the system's loading conditions. To obtain a measured frequency response that's an accurate representation of GH(jω), you must make sure the open-loop system is loaded with the same impedances during testing that will exist when the loop is closed.

You must also avoid system saturation, the condition in which the output at B(jω) gets stuck at the maximum output level when you open the control loop. Such a condition is often caused by the reaction of extremely high loop gains to very small dc offsets—or to integrating components that naturally accumulate stray dc levels—in the absence of the countering effects of negative feedback. When system saturation occurs, you must abandon the loop-open direct technique.

Finally, you must avoid loss of operation (ie, system disruption to the point of failure), which can occur when you open the system's control loop. For example, consider a disk drive's read/write-head-positioning system and the magnetic interface between the head and a prerecorded track on the disk platter. If you open the loop, you won't be able to maintain the interface so that it stays within normal operating conditions. If you stop the disk so that you can position the head over a track on the disk, the magnetic interface will disappear. If you rotate the disk, the slightest off-center condition will cause the head to skip across several tracks. Even the weakest stimulus signal will cause a similar effect. You simply can't test such a system with the loop open.

For a system that's suitable for loop-open direct testing, you should preferably use the DSA's FFT function with a nonperiodic stimulus signal. The FFT function works well because it yields good signal-to-noise performance and because, when you use a nonperiodic stimulus, the DSA allows you to average out distortion products. Further, the FFT function's speed can greatly reduce the total measurement time. When you use a nonperiodic stimulus to reduce the effects of system noise, you must use power-spectrum averaging.

The loop-closed direct method, which is defined by the transfer function Y(jω)/Z(jω), uses the connection shown in **Fig 2**. The technique is a common one for testing systems that can operate stably, though possibly at a reduced performance level, with the control loop closed. You can also use this technique to perform maintenance tests on established control systems, as well as to initiate testing of new systems.

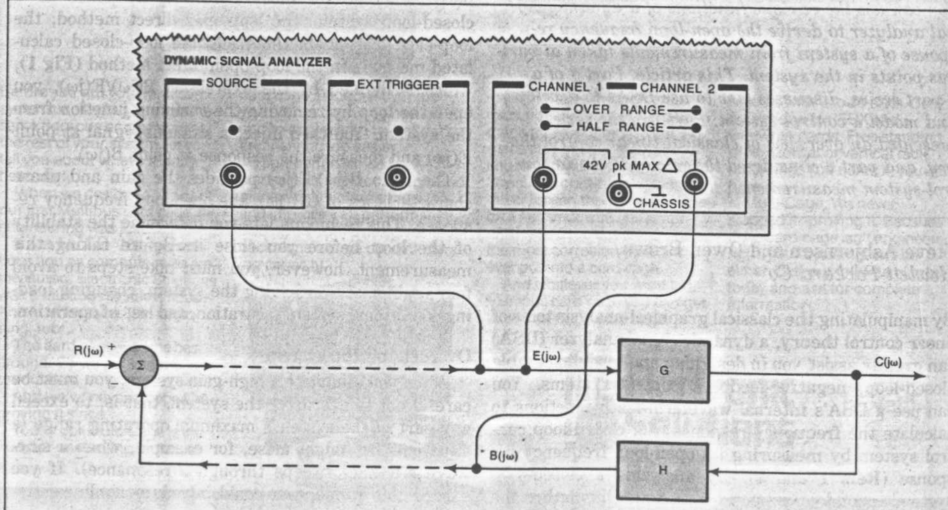

*Fig 1—**To take a loop-open direct measurement,** you remove the summing junction from the closed-loop system. You then inject a stimulus signal at E(jω) and record the system's response at B(jω).*

The measurement technique you choose to derive the open-loop response of your closed-loop system depends on the system's stability and nodal accessibility.

When you test a new system, you should reduce its loop gain to a level that's low enough to ensure loop stability. If the tests verify the system design, you can increase the gain to standard operating levels and then retest the system. To perform the loop-closed direct test, you must monitor the signals that enter and exit any summing junction within the main signal path of the control loop. As long as you don't need to test the control loop in the presence of actual operating signals at the reference input, you can use the negative-feedback summing junction.

When you do need to test the control loop in the presence of actual operating signals at the reference input, or when the system's feedback summing junction is inaccessible (for example, when the head-to-disk magnetic interface of a disk drive is the system's feedback summing junction), you must add a summing junction at some other point in the control loop. Note the summing junction added between the two forward blocks, G_1 and G_2, in the closed-loop system in **Fig 2**. (See **box**, "Create summing junctions.")

In the loop-closed direct method, you inject a stimulus signal into the loop at point S(jω) and monitor the signals at points Y(jω) and Z(jω), the inputs and outputs of the summing junction. In this method, the signal at point Z(jω) is the reference signal. Assuming that all the energy in the loop derives from the stimulus signal S(jω), solving for the signals at Y(jω) and Z(jω) in terms of S(jω) produces the expressions

$$Y(j\omega) = S(j\omega)\frac{G_1G_2H(j\omega)}{1 + G_1G_2H(j\omega)}$$

and

$$Z(j\omega) = S(j\omega)\frac{-1}{1 + G_1G_2H(j\omega)},$$

which is a reasonable assumption to make after you've used averaging to reduce noise. If you take the ratio of Y(jω)/Z(jω), you obtain the equation

$$Y(j\omega)/Z(j\omega) = -G_1G_2H(j\omega),$$

which is the negative of the open-loop frequency response for the system. (The negation, the result of including the negative feedback in the measurement, represents an additional 180° phase shift in the frequency response.) Some test instruments provide you with either waveform-math capabilities or a calibration constant to remove the additional phase shift.

The loop-closed direct technique is a practical one for measuring the frequency response of any system that can operate, at least minimally, when the loop is closed. In such systems, the loop-closed direct method avoids the problems of changing loading conditions, system saturation, and loss of operation that you encounter when you use the loop-open direct method.

Note, however, that when you use the loop-open direct method, you can't use power-spectrum averaging (the standard form of averaging of a DSA's FFT function) to reduce measurement noise. If you attempt to do so, you'll almost always get the wrong frequency-response measurement. Worse, increasing the number of power-spectrum averages will only reduce the variance of the erroneous result, so the measurement will appear to converge to a valid one.

You can't use power-spectrum averaging because any signals within the system not directly related to the stimulus signal (including system noise and any signals applied to the reference input) would pass through the summing junction unaltered. The signals would, therefore, appear in both channels and maintain constant gain and phase relationships (predictably 0 dB and 0°) from measurement to measurement, making it impossible to average them out. And unless you were to compare your measurement with a known-good measurement, you probably wouldn't notice the error.

In any case, a DSA's signal-processing function removes (from each channel) the nonstimulus-related signals in the loop before the DSA calculates the frequency response. The DSA can be operating either in its SFA (swept-frequency analysis) mode or its FFT mode (using time averaging and a periodic stimulus). If coherent distortion products exist in the nonstimulus-related signals, however, the SFA mode is generally the preferred one.

Loop-closed calculated measurement

The transfer function for the loop-closed calculated measurement technique (**Fig 3**) is Y(jω)/S(jω). This method resembles the loop-closed direct method except that here the DSA uses the applied stimulus, instead of the signal at Z(jω), as the reference signal. With the loop-closed calculated method, however, you can use an FFT's power-spectrum averaging mode (and, there-

If a system's feedback summing junction is inaccessible, you can take response measurements by adding a summing junction at another point in the control loop.

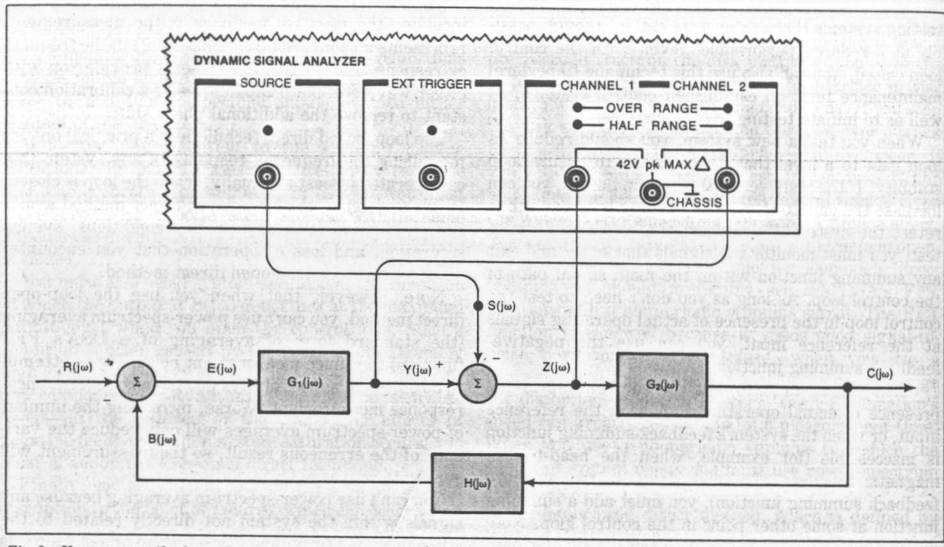

Fig 2—You can use the loop-closed direct measurement method only for systems that are stable in the closed-loop configuration. The technique is useful for negative-feedback systems that need periodic maintenance checks.

fore, you can use random noise) to reject system noise and provide the best linear estimate of a system's frequency response.

You can use random noise because the stray noise in the system appears in only one channel. If you average the relationship between the monitored signals, therefore, eventually the variation (caused by stray noise) in that relationship will drop to zero, as long as the stray noise is not phase-coherent with the stimulus.

The loop-closed calculated technique doesn't limit your selection of an analysis tool. You can use either SFA or FFT analysis with either time or spectrum averaging; for FFT analysis, you can thus use both periodic and nonperiodic stimuli.

The loop-closed calculated method differs from the loop-closed direct method mainly in that, in the loop-closed calculated method, the ratio $Y(j\omega)/S(j\omega)$ does not directly provide the open-loop frequency response, $G_1G_2H(j\omega)$. When you use the equation

$$Y(j\omega) = S(j\omega)\frac{G_1G_2H(j\omega)}{1 + G_1G_2H(j\omega)}$$

to solve for $Y(j\omega)/S(j\omega)$, and then declare the result to be the quantity $T(j\omega)$, you obtain

$$Y(j\omega)/S(j\omega) = \frac{G_1G_2H(j\omega)}{1 + G_1G_2H(j\omega)} = T(j\omega).$$

If you solve for $G_1G_2H(j\omega)$ from this measurement, you obtain

$$G_1G_2H(j\omega) = \frac{T(j\omega)}{1 - T(j\omega)}.$$

Using graphical techniques to perform these calculations would be impractical. DSAs' waveform-math functions, however, perform these calculations automatically. In fact, some of these instruments offer the $T(j\omega)/[1-T(j\omega)]$ calculation as a single-keystroke operation.

The primary disadvantage of the loop-closed calculated technique is that, theoretically, it limits the maximum gain of the open-loop frequency response that you can calculate (from the $Y(j\omega)/S(j\omega)$ measurement) to the dynamic range of one channel of the analyzer. When you measure a signal with a sampling instrument, the instrument's dynamic range is directly

Create summing junctions

To put an electronic summing junction into a control loop, you can use two basic approaches. You can either add new circuitry to the loop to realize the summing junction (Fig A), or you can use an existing buffer amplifier (Fig B). In the first approach, you can calculate the open-loop frequency response from the measured transfer functions $Y(j\omega)/Z(j\omega)$ and $Y(j\omega)/S(j\omega)$:

$$\frac{Y(j\omega)}{Z(j\omega)} = -G_1G_2H(j\omega);$$

$$\frac{Y(j\omega)}{S(j\omega)} = \frac{G_1G_2H(j\omega)}{1 + G_1G_2H(j\omega)} = T(j\omega).$$

Therefore,

$$G_1G_2H(j\omega) = \frac{T(j\omega)}{1 - T(j\omega)}.$$

When you're using an existing amplifier (Fig B) to make a $Y(j\omega)/S(j\omega)$ measurement, you must account for both the gain of the amplifier and the fact that polarities between the $Y(j\omega)$ and $S(j\omega)$ legs of the summing junction now match. The following equations account for these factors:

$$\frac{Y(j\omega)}{Z(j\omega)} = -G_1G_2H(j\omega);$$

$$\frac{Y(j\omega)}{S(j\omega)} = \frac{-G_1G_2H(j\omega)}{\frac{R_1}{R_2} + G_1G_2H(j\omega)} = T(j\omega);$$

$$G_1G_2H(j\omega) = \left(\frac{-T(j\omega)}{1 + T(j\omega)}\right)\left(\frac{R_1}{R_2}\right).$$

In all cases, the amplifiers used to implement the summing junction should have bandwidths much greater than the bandwidth of the control system, and they should also have flat frequency responses within the bandwidth of the system.

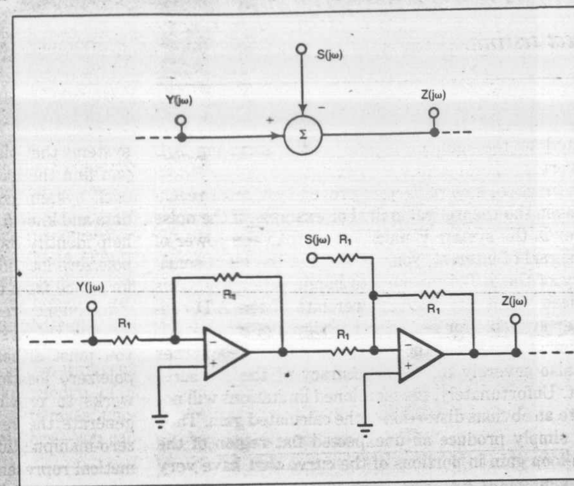

Fig A—This circuit, a typical configuration for adding a summing junction to a control system, allows you to use the measurements of Y(jω)/Z(jω) and Y(jω)/S(jω) in conjunction with a DSA to calculate the system's open-loop frequency response.

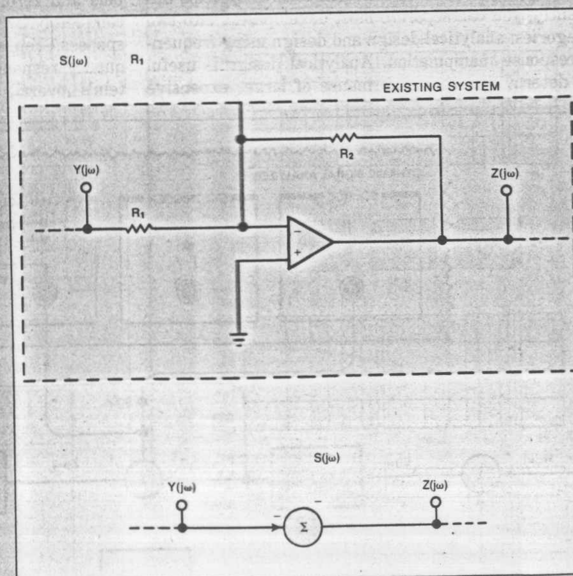

Fig B—You can take advantage of an existing amplifier in a control loop by simply connecting a resistor to the amplifier's summing junction. Using this technique, you take Y(jω)/Z(jω) and Y(jω)/S(jω) measurements and use the DSA's waveform-math function to calculate the system's open-loop frequency response.

Loop-closed direct testing is practical for systems that are stable in closed-loop connection, and it avoids the problems of loop-open direct testing.

related to the number of bits of the sampling A/D converter.

In practice, several factors prevent you from reaching even the theoretical limit. For example, if the noise power in the system is much larger than the power of the signal of interest, you must adjust the input sensitivity of the A/D converter to handle the noise. This reduced sensitivity leaves fewer bits of the A/D converter available for resolving the signal of interest.

Mismatch between the input channels of the analyzer can also severely limit the accuracy of the measurement. Unfortunately, the mentioned limitations will not create an obvious distortion of the calculated gain. They may simply produce an unexpected flat region of the open-loop gain in portions of the curve that have very high loop gains.

Design and modeling

Although you develop different control systems differently, you can separate most development into two categories: analytical design and design using frequency-response manipulation. Analytical design is useful for determining the performance of large, expensive systems that must work the first time. In this method, you find the location of the specific poles and zeros of each system component and use frequency-response data and known properties of the system components to help identify those poles and zeros. Once you know the pole/zero locations, you can calculate the system's performance from the models.

Frequency-response manipulation is useful for improving the system's performance. In this technique, you must either alter the components so that their pole/zero locations change or add compensation networks to provide the poles and zeros necessary to generate the required performance. To use this pole/zero-manipulation process, you must work with mathematical representations of the system and obtain accurate models of the system and its components.

When you add compensation networks, however, you don't need to know the exact location and cause of each pole and zero. Instead, you characterize the system completely by its open- and closed-loop frequency responses. You add compensation networks whose frequency responses will constructively change the system's overall frequency response. In effect, this approach lets you quickly calculate both the frequency response of a compensation network and the network's effect on the system.

The analytical-design and frequency-response-manipulation techniques are similar; the first is simply more concerned with designing with accurate analytical models (typically because of the lack of actual hardware), and the second is more concerned with the measured response of a system.

Typically, you can generate models in two different ways. First, if you have complete knowledge of a device's physical characteristics—such as its mass, friction levels, resistance, and other parameters—you can

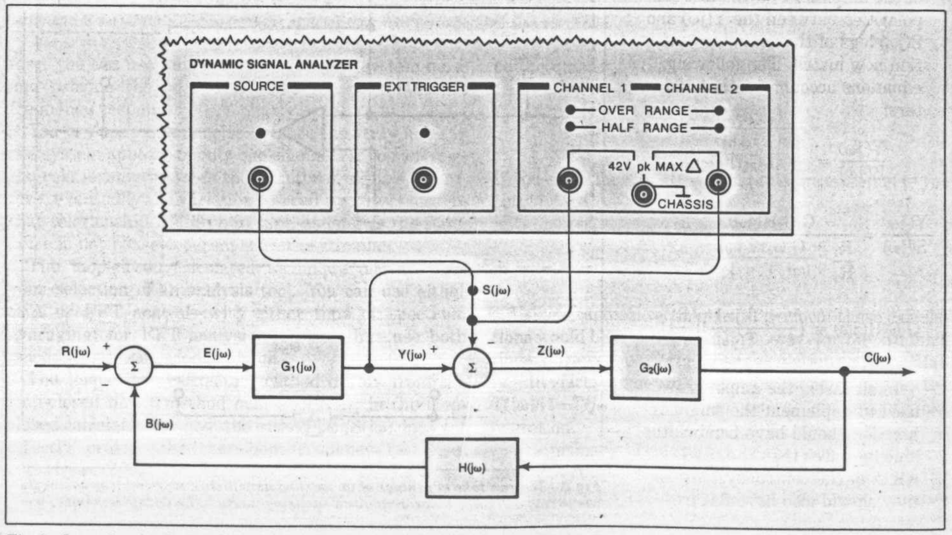

Fig 3—Loop-closed calculated measurement takes advantage of the FFT function's power-spectrum averaging mode. This mode of measurement, when used with a random-noise stimulus, provides the best possible linear estimate of a closed-loop system's frequency response. A DSA's waveform-math capability helps you interpret the results you obtain with this measurement method.

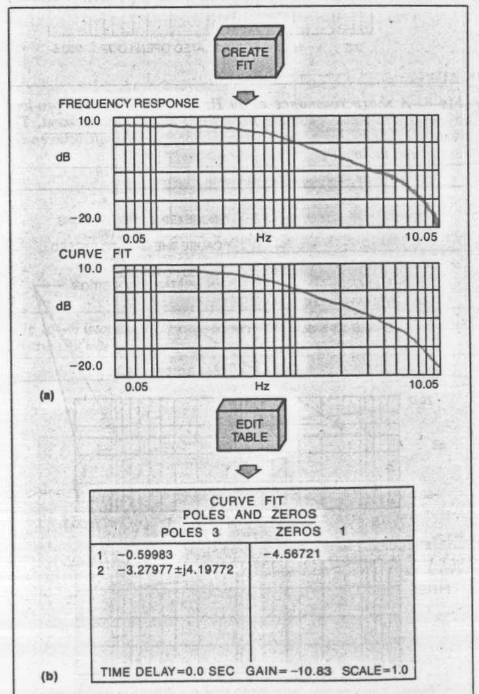

Fig 4—A DSA's curve fitter has an automatic weighting function, which allows the instrument to fit noisy data without producing inconsequential poles and zeros. The DSA displays the system's estimated frequency response below the measured data (a). The edit-table key makes the DSA produce the table of poles and zeros (b) used in the estimate.

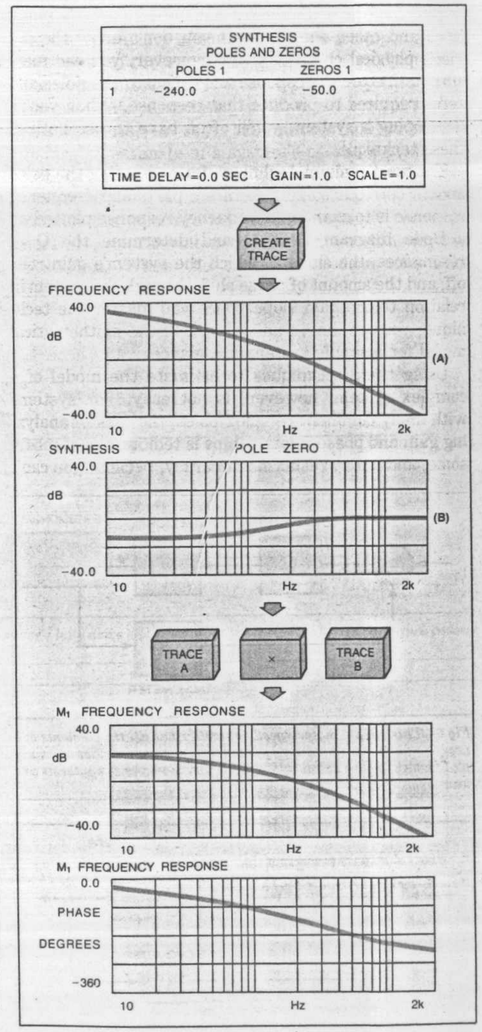

Fig 5—Based on the pole-zero data in a synthesis table, a DSA generates the compensation network's frequency response (trace B). Using waveform math, the DSA multiplies this response by the system's open-loop response (trace A); the results are the gain and phase plots for the compensated system.

In loop-closed calculated measurement, a DSA's FFT function provides a linear estimate of a system's response.

derive a model with them by using Lagrange's equations and other equations. If you don't know the device's physical characteristics, however, you can measure its frequency response and estimate the poles and zeros required to produce that response. When you're developing a system, you'll often have to use both of these techniques.

The most common technique for estimating the poles and zeros required to generate a particular frequency response is to examine a frequency response plotted on a Bode diagram (**Ref 1**) and determine the Q of resonances, the slope at which the system's gain rolls off, and the amount of phase shift through the system in relation to the gain slope. Once you master the technique, you can derive an estimate of the mathematical model for a simple system at a glance.

Using these techniques to estimate the model of a complex system, however, is not easy. For systems with many dominant complex poles, the task of analyzing gain and phase relationships is tedious and cumbersome, and it may result in inaccuracy, because you can't easily sort out the overlapping gain/phase relationships of the many poles and zeros.

Historically, computer programs called curve fitters could automatically analyze a frequency response and produce a pole/zero model. These curve fitters, however, could not distinguish between good data and bad, or noisy, data. If the data was bad, you either had to alter it to remove noise before transferring the data to the computer, or you had to wade through a great deal of inconsequential pole/zero data, which the curve fitter generated to fit the noisy portion of the curve. The computer programs were, therefore, of limited value to designers.

Algorithms called weighting functions improved the usefulness of curve fitters by allowing the computer to analyze the quality of the data before attempting to fit it. Because of these weighting functions, modern curve fitters can successfully process the noisy frequency-response data associated with most control-system measurements, so they can quickly and accurately produce an estimate of a system's pole/zero model.

DSAs contain state-of-the-art curve fitters that are capable of using as many as 40 poles and 40 zeros to estimate the pole/zero model of a frequency response. In a DSA, you can call up the curve fitter with only two keystrokes. When the curve fit is done, the DSA

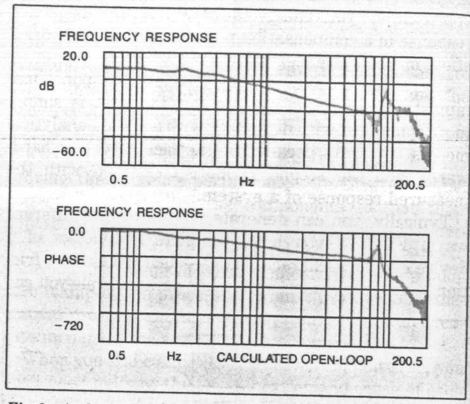

***Fig 8—A sharp resonance at 90 Hz is evident** in this open-loop frequency response obtained from **Fig 7**'s Y/S measurement. The resonance stems from an excessively long drive shaft between the motor and its load.*

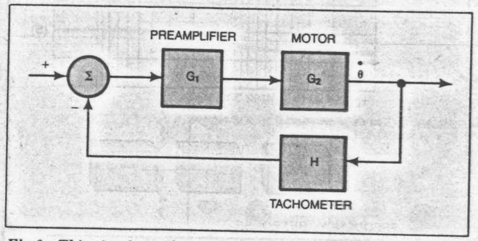

***Fig 6—This simple motor-speed controller** has all the elements of a negative-feedback, closed-loop system. The system provides an example of how you can exploit a DSA's frequency-response synthesis and waveform-math functions.*

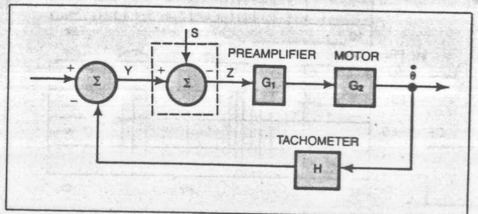

***Fig 7—To perform operational testing at reduced gain** for the normally unstable speed-control system, you add a summing junction at a point in the system before the preamplifier. This test uses the loop-closed calculated technique (**Fig 3**) to obtain a Y/S transfer function.*

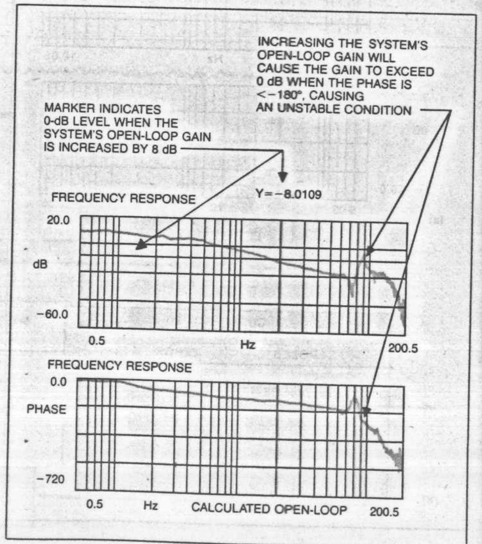

***Fig 9—When you use the DSA's Y-axis marker to show the system's 0-dB level** during normal operation, you can see that the resonance causes the system to become unstable when gain is adjusted to the desired level. For stability, the system's gain must not exceed 0 dB when phase is more negative than −180°.*

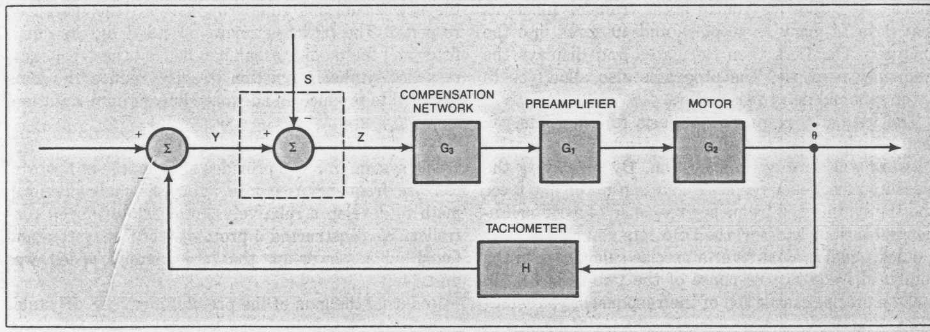

***Fig 10—A lowpass filter counteracts the resonance** arising from the motor's long drive shaft. Without the filter, the 90-Hz resonance causes instability when the system operates at its normal gain setting.*

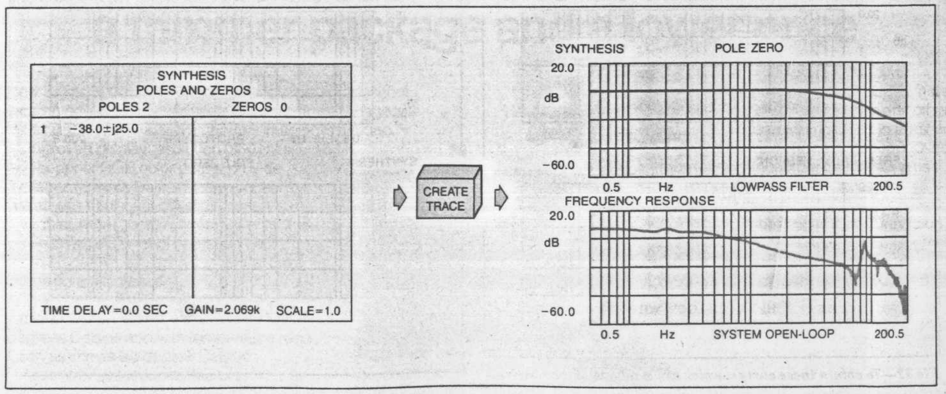

***Fig 11—The DSA's frequency-response-synthesis function** uses pole-zero data and waveform math to create the lowpass filter's response curve, which is displayed above the speed-control system's frequency response.*

A DSA simplifies the task of determining the proper compensation network needed to stabilize closed-loop control systems.

calculates a frequency response from the model and displays that frequency response below the measured frequency response for comparison (**Fig 4a**). Another keystroke displays the table of poles, zeros, and gain (**Fig 4b**) that the curve fit produces.

After you estimate the model of a system or a compensation network, you must calculate the model's frequency response. If you do this calculation manually, you should use Bode's graphical techniques. If you use a DSA, however, you can obtain a higher degree of accuracy for both simple and complex systems.

A DSA's frequency-response-synthesis program performs the inverse function of an advanced dedicated curve fitter, allowing you to enter transfer functions that have as many as 40 poles and 40 zeros into the analyzer. The DSA then calculates and displays the frequency response. The programs also allow you to enter gain and delay parameters.

The frequency-response-synthesis function lets you immediately assess the effect a cascade compensation network will have on your system. By displaying the measured frequency response of a system on one trace and the synthesized frequency response of a compensation network on another, the DSA lets you make either a quick visual evaluation or a precise calculation of the combined frequency response of the two (**Fig 5**). (To make a precise calculation of the responses, you multiply the two frequency responses using the DSA's waveform-math function).

Because curve fitters generate only linear models, the result of the curve fit will be a linear approximation of the device's operation. Should a nonlinearity (other than random noise) produce a large number of poles and zeros, you can reduce the order of the model by simply transferring the curve-fit data to the DSA's frequency-response-synthesis function and selectively subtracting or adding poles and zeros to obtain the best model (with the fewest poles and zeros).

You can then quickly assess the effects of your modifications by synthesizing the frequency response of the new model and comparing it with the measured response. The initial estimate provided by the curve fitter and the modifications handled by the frequency-response-synthesis function greatly reduce the time required to produce an adequate linear representation.

A case study

The system in **Fig 6** provides an example of how you can use frequency-response synthesis and waveform math to develop a relatively simple motor-speed controller. We constructed a prototype of the system and found, upon power-up, that the motor's speed was unstable.

Reducing the gain of the preamplifier by 8 dB stabi-

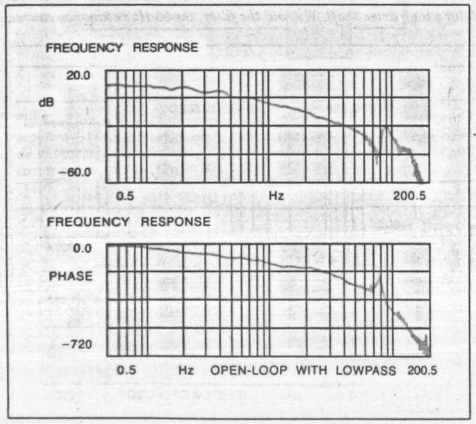

***Fig 12—To obtain these curves,** which are products of the responses of the system and the lowpass filter, you use the DSA's waveform-math function to perform the multiplication. The curves indicate that the use of the filter stabilizes the speed-control system.*

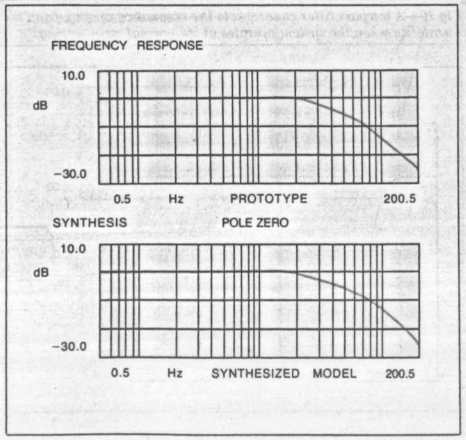

***Fig 13—To check the design** of the lowpass filter, you can compare the filter's measured (upper trace) and theoretical (lower trace) frequency responses.*

The curve-fitter function in a DSA automatically analyzes a system's frequency response and produces a pole/zero plot.

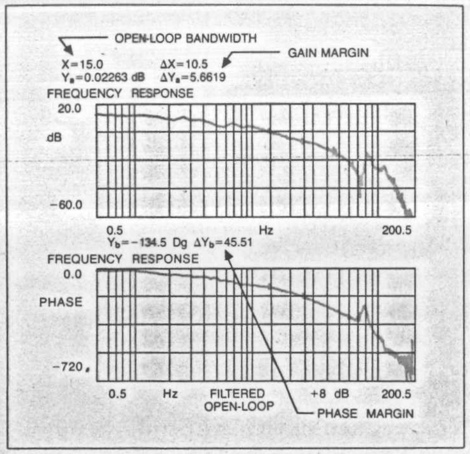

Fig 14—The frequency response of the compensated speed-control system shows that the lowpass-filter compensation network completely stabilizes the system. The DSA gives direct readouts of gain margin and phase margin.

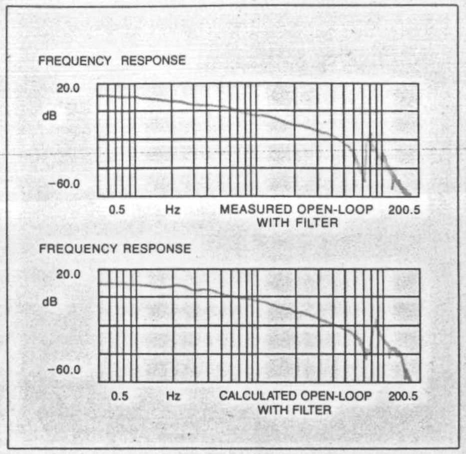

Fig 15—A final check on the speed-control system shows that the system's measured frequency response (upper trace) closely matches the response predicted (lower trace) by the DSA's frequency-response synthesis and waveform-math function.

lized the system, but degraded its performance below acceptable levels. While the system was at the reduced-gain level, we measured the open-loop frequency response by adding a summing junction to the system just before the preamplifier (Fig 7) and by using the loop-closed calculated (Y/S) measurement technique. Fig 8 shows the open-loop frequency response calculated from the Y/S measurement.

The measurement revealed a sharp resonance at approximately 90 Hz. When we placed the Y-axis marker at −8 dB to indicate the unity-gain (0-dB) location during normal system operation (Fig 9), we saw that the resonance did indeed cause the system to become unstable when gain was adjusted to the desired level.

The source of the resonance was a relatively long drive shaft attached between the motor and the anticipated system load. Redesigning the system would be too expensive, so we decided not to alter the drive shaft, but to find an electronic solution.

Because the problem was occurring at a relatively high frequency, we added compensation to the system in the form of a lowpass filter (Fig 10). Using the DSA's frequency-response-synthesis function, we entered the pole locations for a simple lowpass filter into the analyzer and synthesized and displayed the frequency response above the system's measured frequency response (Fig 11).

By using its waveform-math function to multiply the two frequency responses, the DSA predicted the effect of adding the lowpass filter to the system's loop. Fig 12 shows the gain and phase of the modified system. The predicted response of the modified system indicated that the lowpass filter would be a good solution to the resonance problem, so we constructed a prototype of the filter.

To ensure that the design was correct, we measured and compared the frequency response of the filter prototype to its synthesized frequency response (Fig 13). Once we had confirmed its design, we added the filter to the system and increased the preamplifier's gain by 8 dB to its previous level. The system remained stable, but we measured the open-loop frequency response again to make sure that the gain and phase margins were sufficient to maintain stability.

The measurement indicated that the resonance had indeed been attenuated far below any level of concern and, using the analyzer's markers, we quickly recorded the gain margin, phase margin, and open-loop bandwidth (Fig 14). To evaluate the integrity of the design approach, we compared the measured open-loop frequency response of the modified system with the pre-

dicted response (Fig 15).

As this example demonstrates, the DSA's design tools let you compare predicted and measured responses almost effortlessly, without using data transfers or plotted data. The DSA's advanced measurement and analysis tools provide a key link between test, analysis, model, and design. **EDN**

References

1. Asbjornsen, Steve, and Brown, Owen, "Apply control-system theory to analyze closed-loop systems," *EDN*, April 3, 1986, pg 173.
2. Asbjornsen, Steve, and Brown, Owen, "Dynamic signal analyzers simplify measurement of linear control systems," *EDN*, April 17, 1986, pg 219.

Authors' biographies

Steven W S Asbjornsen is a product-marketing engineer at Hewlett-Packard (Everett, WA), where he's responsible for control-system application and market development. He received a BSEE from the University of Washington and has been with HP for five years. Steve, who is a musician in his spare time, also enjoys hiking and waterskiing.

Owen Brown, who is in charge of product publicity for Hewlett-Packard (Palo Alto, CA), has worked for HP for two years. He has a bachelor's degree from Yale University and an MBA from the University of Chicago. His spare-time pursuits include painting and aikido.

Article Interest Quotient (Circle One)
High 485 Medium 486 Low 487